Opinion

Cell PRESS

# New tricks with old genes: the genetic bases of novel cnidarian traits

Sylvain Forêt[1,2], Brent Knack[1], Evelyn Houliston[3], Tsuyoshi Momose[3], Michael Manuel[4], Eric Quéinnec[4], David C. Hayward[2], Eldon E. Ball[2] and David J. Miller[1]

[1] ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland 4811, Australia
[2] Centre for the Molecular Genetics of Development, Australian National University, PO Box 475, Canberra ACT 2601, Australia
[3] UPMC Université Paris 06 and Centre National de la Recherche Scientifique UMR 7009 ''Biologie du Développement'', Observatoire Océanologique, 06230 Villefranche-sur-Mer, France
[4] UMR 7138 Systématique, Adaptation, Evolution, Université Pierre et Marie Curie, 7 quai St Bernard, 75005 Paris, France

**Recent thought on genome evolution has focused on the creation of new genes and changes in regulatory mechanisms while ignoring the role of selective gene loss in shaping genomes. Using data from two cnidarians, the jellyfish *Clytia* and the coral *Acropora*, we examined the relative significance of new 'taxonomically restricted' genes and selectively retained ancestral genes in enabling the evolution of novel traits. Consistent with its more complex life-cycle, the proportion of novel genes identified in *Clytia* was higher than that in the 'polyp only' cnidarians *Nematostella* and *Hydra*, but each of these cnidarians has retained a proportion of ancestral genes not present in the other two. The ubiquity and near-stochastic nature of gene loss can explain the discord between patterns of gene distribution and taxonomy.**

## The genetic basis of new traits

The evolution of new traits and new characteristics requires changes in the genome. The role of changes in regulatory mechanisms is well documented, but studies of cnidarians in particular have highlighted the evolutionary significance of changes in the proteome. Whilst changes in regulatory mechanisms and the co-option of existing genes to new roles are undoubtedly important in enabling evolutionary novelties, 'new' genes also clearly contribute. New genes arise constantly via duplication events, and can diverge rapidly to the extent that their origin becomes unclear on a relatively short time scale [1]. Many of the taxonomically restricted genes (TRGs) that are presently under scrutiny in several laboratories (see below) probably arose in this way. Other mechanisms suggested to account for new genes include lateral (or horizontal) gene transfer (LGT; see Ref. [2] for a possible example from cnidarians), although this is controversial with respect to the animal kingdom (e.g. see Ref. [3]), and genes originating *de novo* from non-coding sequences. Here, we focus on a distinct mechanism for the generation of TRGs, which has received little attention until now, the selective retention of ancestral genes in the face of generalized gene loss across other lineages. This idea acknowledges the accumulating evidence that the ureumetazoan (the common ancestor of the eumetazoan clade comprising the Cnidaria and Bilateria) was genetically complex, and that gene losses have been ubiquitous across the animal kingdom. If all animals have undergone gene loss, then retention of more of the ancestral gene set by some could potentially contribute to the evolution of new traits [4]; effectively, fewer losses might translate to improved 'evolvability' in some lineages. This leads to the prediction that some of the genes responsible for the evolution of novel traits might have homologs in distantly related animals but not in close relatives.

## New genes for new traits

Although the basic idea is not new, a flurry of recent papers has led to renewed interest in the idea that TRGs enable novel traits. Thomas Bosch's group has published several elegant papers linking *Hydra* TRGs to new traits (reviewed in Ref. [5]). Published analysis of a fully sequenced genome typically reports a surprisingly high proportion of 'novel' or 'unique' genes; however, variations in these numbers reflect arbitrary cutoffs as well as the state of the database at the time of the analysis, and the sensitivity of the algorithms used. Thus, a proportion of genes reported as unique appear so only because of the limited range of animals for which whole genome sequences were available at the time of the analysis. Although some groups (e.g. chordates) are reasonably well represented in sequence databases, the genomes of much of the animal kingdom remain uncharted waters. Consequently, current estimates of novel genes are overestimates and subject to ongoing downward review. There are many examples of genes identified as 'taxonomically restricted' that have proven to be more widely distributed; for example, proteins related to jellyfish green fluorescent protein were thought to be restricted to the Cnidaria but have recently been identified in amphioxus [6] and in copepods [7]. Moreover, of those putative novel genes, many are probably highly diverged members of known gene families. For example, the human male sex determinant *Sry* is a highly diverged Sox gene, whose most likely origin was as a paralog of *Sox3*, but which is restricted to eutherian mammals [8].

## New traits within phylum Cnidaria

In spite of their simplicity in terms of cell type diversity, cnidarians belong to one of the most diverse and successful

*Corresponding author:* Miller, D.J. (david.miller@jcu.edu.au).

0168-9525/$ – see front matter  doi:10.1016/j.tig.2010.01.003 Available online xxxxx

animal phyla, with members inhabiting almost every possible aquatic habitat, facilitated by the evolution of a similarly diverse range of new traits. Four classes are traditionally recognized within the Cnidaria. Of these, the Anthozoa (which includes corals and sea anemones) have the simplest life-cycles with only a single mature morphology, the polyp (Figure 1). Innovations in the class Anthozoa include the coupled phenomena of massive calcification, metamorphosis and symbiosis in the stony corals (Scleractinia). One would expect to see novel genes, in this case genes not present in the sea anemone, involved in these derived processes. The Medusozoa, including the other three cnidarian classes Scyphozoa, Cubozoa and Hydrozoa (Figure 1), typically go through a polyp stage and a medusa (jellyfish) stage, in their life-cycle. The medusa is anatomically much more sophisticated than any anthozoan polyp. The musculature is more advanced (e.g. the presence of striated muscle), the nervous system is more organized and heterogeneous, and various complex sense organs are known. For instance, cubozoans (box jellyfish) have evolved complex eyes and fast swimming behavior, enabling them to be aggressive hunters. Molecular phylogenies firmly and unambiguously place the Medusozoa as a monophyletic sister group to the Anthozoa; so, unless loss of the medusa occurred somewhere during anthozoan evolution, the polyp-only life-cycle reflects the ancestral condition. Under the 'new genes for new traits' scenario, the numerous morphological innovations of medusozoans are thus generally predicted to require the evolution of novel genes.

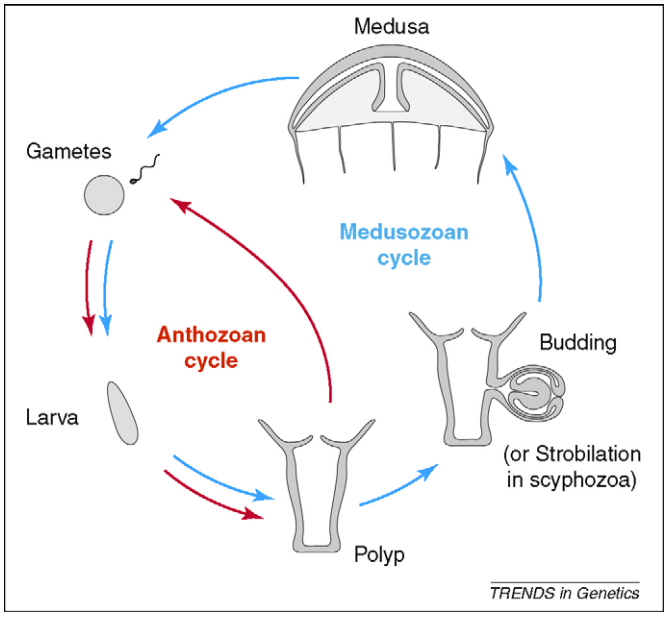


**Figure 1.** Cnidarian life-cycles. Most cnidarians have a planula larva that metamorphoses into a polyp, which is often sessile. In anthozoans the polyp is the sexual stage; however, members of the three other cnidarian classes (collectively known as the Medusozoa) reproduce sexually as free-swimming medusae (jellyfish), derived asexually from the polyp. Medusae are generally more complex than the polyps of the same species or those of anthozoans. Medusae arise in a wide variety of ways, including budding (Hydrozoa), strobilation (Scyphozoa) and metamorphosis (Cubozoa). In colonial hydrozoans, including *Podocoryne* and *Clytia*, specialized polyps known as gonozoids are dedicated to budding. In the solitary hydrozoan *Hydra*, the medusa and planula phases have been secondarily lost, whereas in many trachyline hydrozoans the polyp phase has been lost or greatly reduced. The polyp-only phase present in class Anthozoa is generally believed to be the ancestral state.

## New genes in the biphasic hydrozoan *Clytia*, a 'typical' medusozoan

Whole genome sequence data are currently available for two cnidarians, the anthozoan *Nematostella* and the hydrozoan *Hydra*, a medusozoan that has secondarily lost the medusa life-cycle phase. Although whole genome sequences from a wider range of cnidarians will be required for a definitive evolutionary picture to emerge, a large transcriptome dataset is now available for the hydrozoan *Clytia hemisphaerica*, and this can provide some initial insights into the gene repertoire of a more typical medusozoan with both polyp and medusa phases in its life-cycle. The publicly available *Clytia* EST data (~ 80000 ESTs) corresponds to just over 8000 unigenes. Consistent with the idea that the evolution of new traits requires novel genes, simple reciprocal Blast comparison (see the supplementary material online) shows that the *Clytia* dataset contains a higher proportion of unique sequences than those of either *Hydra* or *Nematostella*; the proportion of the 8219 unigenes in *Clytia* with no significant matches is around 18% (22% if *Hydra* is excluded from the analysis), whereas the corresponding number is 11% for both *Hydra* and *Nematostella* (Figure 2a). Confirmation of the proportion of unique sequences awaits the forthcoming whole genome sequence but, although such a simple-minded approach is easy to criticize, the substantial difference observed suggests a real effect. As might be expected, more sequences are hydrozoan-specific (i.e. present only in *Hydra* and *Clytia*) than are shared by *Nematostella* and either or both hydrozoans. It is noteworthy, however, that each of these three animals, irrespective of its morphological complexity, has over 10% new sequences. In *Clytia*, the medusa is not the only innovation relative to the ancestral condition; new genes might also be required to enable, for example, colonial organization and polyp dimorphism, whereas in the case of *Hydra* some new genes have presumably been required to enable life in a fresh-water environment. Interestingly, contrary to the common view that sea anemones are more "ancestral" and hydras are more "derived", they have similar percentages of unique sequences.

One example of a hydrozoan-restricted gene family is that encoding proteins containing the sweet tooth domain, a novel family of receptor tyrosine kinases implicated in carbohydrate binding [9]. Amongst the genes identified as unique to *Clytia* are a novel family of seven transmembrane domain (7TM) proteins (Forêt *et al*. unpublished results). Several clade-specific 7TM protein families have been described, such as the vertebrate, insect or nematode olfactory receptors, which do not share any sequence similarity [10]. At first glance, these would qualify as TRGs; however, rather than the same architecture having been invented independently in these clades, it is more likely that these sequences have a common origin and have evolved beyond recognition as a result of relaxed or positive selection.

## The family jewels; ancestral genetic complexity

Analyses of EST [11,12] and genomic [13] data for the anthozoan cnidarians *Nematostella* and *Acropora* have provided strong support for the idea of a genetically com-

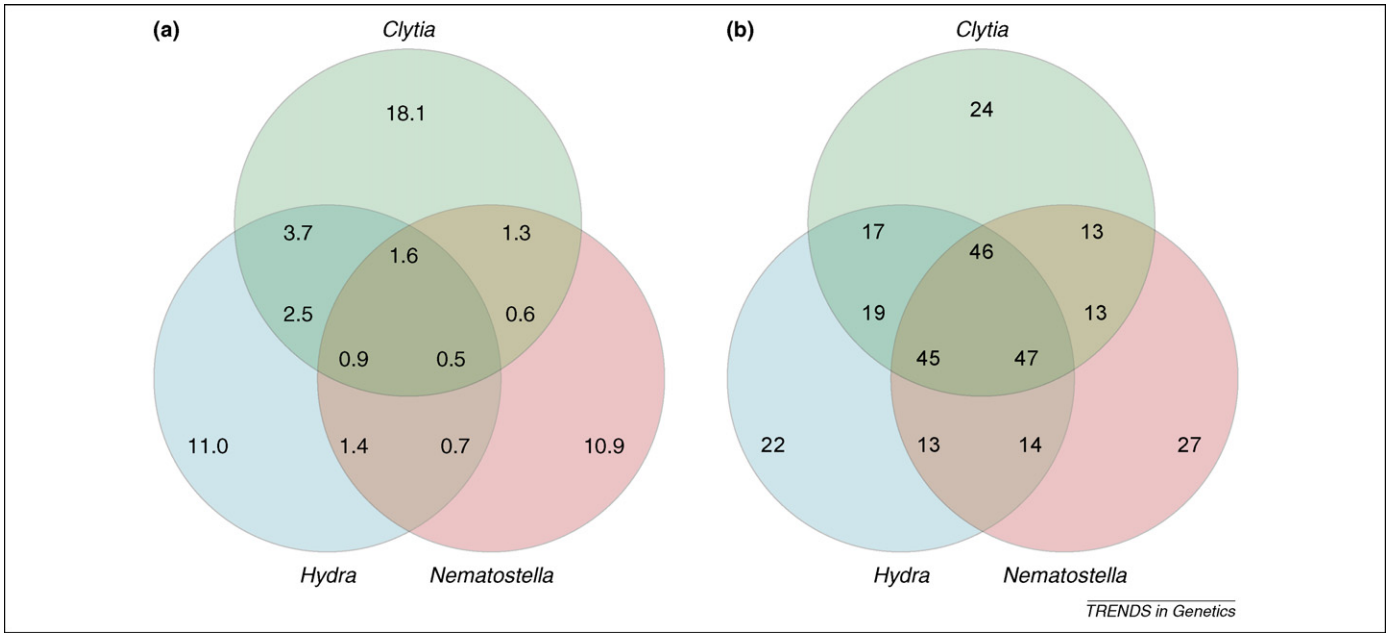


**Figure 2**. Comparisons of the (incomplete) *Clytia* EST dataset with the predicted proteomes of two polyp-only cnidarians. (**a**) Distribution of cnidarian-specific sequences in three cnidarians. The numbers are expressed as percentages of the number of protein-coding genes available (8053, 17398 and 24780 for *Clytia, Hydra* and *Nematostella,* respectively); where a sector of the Venn diagram contains more than one number, each refers to the percentage of sequences in the proximal species. Sequences with no homolog in RefSeq beside *Hydra* and *Nematostella* were classified as cnidarian specific; a BLAST e-value cutoff of $1e^{-5}$ was used for homology calls. The numbers of unigenes that have no non-cnidarian match are 2038 in *Clytia,* 2764 in *Hydra* and 3219 in *Nematostella*. (**b**) Distribution of cnidarian sequences with significant RefSeq matches. To compensate for biases potentially arising due to the smaller available *Clytia* dataset, the numbers of sequences uniquely shared by one cnidarian and RefSeq were calculated on the basis of resampling from the *Clytia, Hydra* and *Nematostella* datasets. The numbers presented here are the average of 1000 samples of 6000 genes (for details of the method, see the supplementary material online). With the caveat that these estimates are based on resampling, they imply extensive gene loss in each of the three species considered. Definitive values, however, await the availability of the *Clytia* genome sequence.

plex common eumetazoan ancestor, as have studies of specific gene families. For example, *Nematostella* has a near-complete Wnt complement [14]; members of 12 of the 13 recognized vertebrate and invertebrate Wnt subfamilies. These studies have led to the assumption that *Nematostella* has faithfully maintained the ancestral condition. On the other hand, *Hydra* is often regarded as a highly derived cnidarian and has clearly undergone significant secondary gene loss. For example, whereas *Nematostella* has unambiguous homologs of both the Toll receptor and NFκb, *Hydra* has neither [15]. In addition, *Hydra* has a much-reduced complement of homeobox genes [16] and is missing two Wnt subfamilies that are represented in *Nematostella* [17]. The true extent of gene loss in *Hydra* remains to be determined.

Although quantitative analysis is not yet appropriate, the EST data available for *Clytia* should allow a more informative comparison with *Nematostella* with respect to retention of the ancestral gene complement. Analysis of the cnidarian complements of the Sox [18] and Fox [19] transcription factor families have already revealed a wide representation of each family amongst the *Clytia* ESTs, with examples of specific losses in *Clytia* and *Nematostella*. To address this question more globally, the numbers of *Hydra*, *Nematostella* and *Clytia* protein sequences with significant RefSeq matches were compiled, and each of the cnidarian datasets was scanned for their presence or absence (Figure 2b). Whilst this analysis is subject to the caveat that the *Clytia* data are incomplete, two implications of this are, first, that extensive gene losses have occurred in each of the cnidarian species examined, while other ancestral genes have been uniquely retained. Second, *Nematostella* has not maintained dramatically more ancestral sequences even than *Hydra*. The relatively small difference implied by this analysis contrasts with the common perception of *Nematostella* as having maintained far more of the ancestral gene set than *Hydra*. Our results, however, are not incompatible with the reduction of some gene families in *Hydra*.

## Stochastic gene loss leads to unpredictable gene distributions

Amongst the *Clytia* genes with RefSeq matches that are not present in *Nematostella* or *Hydra* (Figure 2b) are clear homologs of the CELIII lectins that were previously known only from the sea cucumber *Cucumaria* (an echinoderm) and *Acropora* [20]. As in *Acropora*, there are two clear CELIII matches in the *Clytia* EST dataset (Knack *et al*., unpublished results). CELIII is an atypical B-type lectin [21,22] with two carbohydrate-binding domains that adopt ricin-like β-trefoil folds [23], and a C-terminal oligomerization domain. This group of proteins is unique, the *Cucumaria*, *Acropora* and *Clytia* sequences have high overall primary sequence similarity but are clearly distinct from all other proteins in Genbank. At the tertiary structure level there are similarities between the N-terminal domains of CELIII with ricin and other members of the (QXW3) motif family, and the overall level of similarity between the cnidarian CELIII-like proteins and the *Cucumaria* CELIII implies that their 3D structures are likely to be very similar. The *Clytia* and *Acropora* genes are the products of independent duplication events (data not shown), and expression data are not yet available for the former. The distribution of CELIII-like lectins strikingly

illustrates both the ubiquity of gene loss and its apparently stochastic nature; parallel losses have occurred in the cnidarian lineages Hydrozoa and Anthozoa, and apparently also on multiple occasions elsewhere in the animal kingdom.

### Both old and new genes enable novel traits in corals

Recent work on coral metamorphosis and calcification implicates both novel (coral-specific) and ancestral genes in enabling these derived traits [20,24–26]. The ability to secrete a calcium carbonate skeleton has evolved independently many times in the animal kingdom, so consequently this trait appears to involve unique genes (TRGs) in different lineages. Modern reef-building (scleractinian) corals build massive skeletons of calcium carbonate in the form of aragonite, and first appear in the fossil record in the mid Triassic. By comparison with other calcifying phyla, coral calcification is poorly understood at the molecular level, but some recent progress has been made. The galaxin genes are an example of a family of TRGs originally thought to be restricted to reef-building corals [24,25]. Galaxins were identified as major protein components of the organic matrix that controls the deposition of calcium carbonate in the coral *Galaxea* [24]. Subsequently, three galaxin-related proteins have been identified in *Acropora* and the expression patterns of these are consistent with roles in calcification [25]. Galaxins were originally reported as unique to corals, however, a 'galaxin' has been reported in the tubeworm *Riftia* [27] and there are database entries annotated as "similar to galaxin" (*Ciona* GI:198432839, *Caenorhabditis elegans* GI:31746535) or "galaxin-related" (*Oikopleura* GI:42601315). Whilst clear orthologs of galaxin may be restricted to scleractinians, the characteristics of these more distant homologs suggest that the precursors of the galaxins were recruited (most likely from an ancient role in the extracellular matrix) to roles in coral calcification during the Triassic, and have subsequently diverged to the point where their origins are obscured and they might now appear to be TRGs. Sunigawa *et al*. [26] recently reported the identification of a family of small, cysteine-rich proteins (SCRiPs) that appear to be restricted to scleractinians, as no clear homologs are present in *Nematostella* or elsewhere. Microarray experiments [20] suggest a variety of possible roles for SCRiPs, but the timing of expression of some of these TRGs is again consistent with a role in calcification.

Although some novel genes have been implicated in calcification in corals, the linked trait of complex metamorphosis provides a number of examples of selective and specific retention of ancestral genes. All anthozoans have a motile planula larva stage but, whereas sea anemones such as *Nematostella* [28] display a very 'simple' and non-selective metamorphosis from motile larva to relatively sessile polyp, in many corals that process is selective, discrete and much more dramatic. Part of the complex metamorphosis seen in corals might be a consequence of the need to transform the aboral ectoderm into calicoblastic epithelium, the tissue responsible for secretion of the skeleton. As the coral version of metamorphosis is presumably a derived trait, one might expect TRGs to again be key players. However, several genes with clear counterparts in other animals yet apparently absent from *Nematostella* are heavily and specifically expressed during metamorphosis in the coral *Acropora*. These include homologs of the perforin domain apextrin [15] and the haemolytic lectin CELIII [20], both presumably components of the innate immune repertoire and both previously known only from (different) echinoderms (note that apextrin was first cloned as a TRG in a direct developing sea urchin [29]). Thus, genes encoding apextrin and the CELIII lectins appear to be retained from the common ancestor by *Acropora* but not by *Nematostella*, and are possibly involved in the more complex metamorphosis in the former.

### Concluding remarks

To summarize, three general conclusions follow from the arguments presented here. First, although it has sometimes been assumed that anthozoan cnidarians have essentially frozen the complex ancestral gene set, acquisition of new genes and extensive losses clearly have occurred similarly in both anthozoans and hydrozoans. The true extent of gene loss in various lineages will be clear only when many more whole genome sequences are available. Second, whilst the evolution of new and taxon-specific traits undoubtedly involves some new genes (TRGs), genes retained from the common ancestor but lost elsewhere might also contribute to novel characteristics. Third, the ubiquity and apparently near-random nature of gene loss across the Metazoa can generate patterns of gene distribution that appear inconsistent with phylogeny. Such patterns can be misinterpreted as reflecting lateral transfer events.

#### Acknowledgements

We thank the Genoscope (Evry, France) for generating the *Clytia* transcriptome data, and the Australian Research Council for support.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2010.01.003.

#### References

1 Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
2 Denker, E. *et al.* (2008) Horizontal gene transfer and the evolution of cnidarian stinging cells. *Curr. Biol.* 18, R858–R859
3 Stanhope, M.J. *et al.* (2001) Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* 411, 940–944
4 De Robertis, E.M. (2008) Evo-devo: variations on ancestral themes. *Cell* 132, 185–195
5 Khalturin, K. *et al.* (2009) More than just orphans: are taxonomically-restricted genes important in evolution. *Trends Genet.* 25, 404–413
6 Deheyn, D.D. *et al.* (2007) Endogenous green fluorescent protein (GFP) in amphioxus. *Biol. Bull.* 213, 95–100
7 Shagin, D.A. *et al.* (2004) GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Mol. Biol. Evol.* 21, 841–850
8 Grützner, F. *et al.* (2004) In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature.* 432, 913–917
9 Reidling, J.C. *et al.* (2000) Sweet tooth, a novel receptor protein-tyrosine kinase with extracellular C-type lectin-like domains. *J. Biol. Chem.* 275, 10323–10330

10 Mombaerts, P. (1999) Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* 286, 707–711
11 Kortschak, R.D. *et al.* (2003) EST analysis of the cnidarian, Acropora millepora, reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Curr. Biol.* 13, 2190–2195
12 Technau, U. *et al.* (2005) Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends Genet.* 21, 633–639
13 Putnam, N.H. *et al.* (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science.* 317, 86–94
14 Kusserow, A. *et al.* (2005) Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* 433, 156–160
15 Miller, D.J. *et al.* (2007) The innate immune repertoire in Cnidaria – ancestral complexity and stochastic gene loss. *Genome Biol.* 8, R59
16 Chourrout, D. *et al.* (2006) Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature* 442, 684–687
17 Lengfeld, T. *et al.* (2009) Multiple Wnts are involved in Hydra organizer formation and regeneration. *Dev. Biol.* 330, 186–199
18 Jager, M. *et al.* (2006) Expansion of the Sox gene family predated the emergence of the Bilateria. *Mol. Phylogenet. Evol.* 39, 468–477
19 Chevalier, S. *et al.* (2006) Polarised expression of FoxB and FoxQ2 genes during development of the hydrozoan *Clytia hemisphaerica*. *Dev. Genes Evol.* 216, 709–720
20 Grasso, L. *et al.* (2008) Microarray analysis provides candidate genes for key roles in coral development. *BMC Genomics* 9, 540
21 Nakano, M. *et al.* (1999) Primary structure of hemolytic lectin CEL-III from marine invertebrate *Cucumaria echinata* and its cDNA: structural similarity to the B-chain from plant lectin, ricin. *Biochim. Biophys. Acta* 1435, 167–176
22 Kouzuma, Y. *et al.* (2003) Characterization of functional domains of the hemolytic lectin CEL-III from the marine invertebrate *Cucumaria echinata*. *J. Biochem.* 134, 395–402
23 Uchida, T. *et al.* (2004) Crystal structure of the hemolytic lectin CEL-III isolated from the marine invertebrate *Cucumaria echinata*: implications of domain structure for its membrane pore-formation mechanism. *J. Biol. Chem.* 279, 37133–37141
24 Fukuda, I. *et al.* (2003) Molecular cloning of a cDNA encoding a soluble protein in the coral exoskeleton. *Biochem. Biophys. Res. Commun.* 304, 11–17
25 Reyes-Bermudez, A. *et al.* (2009) Differential expression of three galaxin-related genes during settlement and metamorphosis in the scleractinian coral *Acropora millepora*. *BMC Evol. Biol.* 9, 178
26 Sunagawa, S. *et al.* (2009) Identification and gene expression analysis of a taxonomically restricted cysteine-rich protein family in reef-building corals. *PLoS One* 4, e4865
27 Sanchez, S. *et al.* (2007) Identification of proteins involved in the functioning of *Riftia pachyptila* symbiosis by subtractive suppression hybridization. *BMC Genomics* 8, 337
28 Daly, M. *et al.* (2008) Phylogenetic relationships among sea anemones (Cnidaria: Anthozoa: Actiniaria). *Mol. Phylogenet. Evol.* 48, 292–301
29 Haag, E.S. *et al.* (1999) Apextrin, a novel extracellular protein associated with larval ectoderm evolution in *Heliocidaris erythrogramma*. *Dev. Biol.* 211, 77–87